We claim:

1. A computer-implemented method comprising:

allocating each of a plurality of items to at least one of a plurality of clusters, based on a predetermined criterion accounting for at least a quota for each item;

selecting an item for a current cluster from items allocated to the current cluster; and,

effecting the item.

2. The method of claim 1, wherein the plurality of items comprises a plurality of ads, and effecting the item comprises displaying the ad.

3. The method of claim 2, wherein the predetermined criterion further accounts for a constraint for each cluster.

4. The method of claim 2, wherein the predetermined criterion further accounts for a particular one of the plurality of ads restricted from being shown in a particular one or more of the plurality of clusters.

5. The method of claim 2, wherein the predetermined criterion comprises maximizing an expression $\sum_{ij} p_{ij} x_{ij}$, where p_{ij} comprises a probability that a user in cluster j will actuate ad i.

6. The method of claim 5, wherein the predetermined criterion further comprises maximizing the expression subject to a constraint $\sum_j x_{ij} = q_i$, where q_i comprises a quota for ad i, and x_{ij} comprises a total number of times ad i is shown in cluster j.

7. The method of claim 5, wherein the predetermined criterion further comprises maximizing the expression subject to a constraint $\sum_i x_{ij} = c_j$, where c_j comprises a constraint for cluster j, and x_{ij} comprises a total number of times ad i is shown in cluster j.

8. The method of claim 5, wherein the predetermined criterion comprises maximizing the expression subject to a first constraint $\sum_j x_{ij} = q_i$, where q_i comprises a quota for ad i, and x_{ij} comprises a total number of times ad i is shown in cluster j, and a second constraint $\sum_i x_{ij} = c_j$, where c_j comprises a constraint for cluster j, and x_{ij} comprises a total number of times ad i is shown in cluster j, such that the expression, the first constraint and the second constraint define a linear program.

9. The method of claim 8, wherein the linear program is solved by the Simplex Algorithm.

10. The method of claim 2, wherein allocating each of a plurality of ads to at least one of the plurality of clusters comprises determining for each ad in each cluster a probability that a user in the cluster will actuate the ad.

11. The method of claim 10, wherein the probability that a user in the cluster will actuate the ad comprises the probability that a user in the cluster will click on the ad.

12. The method of claim 10, wherein determining for each ad in each cluster a probability that a user in the cluster will actuate the ad comprises inputting training data from which to determine for each ad in each cluster the probability that a user in the cluster will actuate the ad.

13. The method of claim 10, wherein determining for each ad in each cluster a probability that a user in the cluster will actuate the ad comprises utilizing at least one of: a maximum likelihood approach, a MAP method approach, and, a hierarchical Bayesian approach.

14. The method of claim 2, wherein the predetermined criterion comprises maximizing an expected number of actuations of the plurality of ads, given the quota for each ad and the constraint for each cluster.

15. The method of claim 2, wherein the constraint for each cluster comprises a total number of times the cluster is visited by any user.

16. The method of claim 2, wherein the quota for each ad comprises a total number of times that the ad must be displayed.

17. The method of claim 2, wherein the criterion comprises favoring at least one ad over other ads within the plurality of ads in allocating the at least one ad.

18. The method of claim 2, wherein the criterion comprises accounting for at least one house ad.

19. The method of claim 2, wherein the predetermined criterion comprises minimizing an expression $\sum_{ij} p_{ij}x_{ij}$, where p_{ij} comprises a probability that a user in cluster j will actuate ad i.

20. The method of claim 2, wherein the predetermined criterion comprises maximizing an expression $\sum_{ij} \alpha_i p_{ij}x_{ij}$, where p_{ij} comprises a probability that a user in cluster j will actuate ad i, and α_i comprises a coefficient for the ad i to indicate weighting of the ad i.

21. The method of claim 5, wherein the predetermined criterion further comprises maximizing the expression subject to a constraint $x_{ij}=0$ for a particular ad i within a particular cluster j, where x_{ij} comprises a total number of times the ad i is shown in the cluster j.

22. The method of claim 5, wherein the predetermined criterion further comprises maximizing the expression subject to a constraint $\sum_i x_{ij} \leq c_j$, where c_j comprises a constraint for cluster j, and x_{ij} comprises a total number of times ad i is shown in cluster j.

23. The method of claim 10, wherein the probability that a user in the cluster will actuate the ad comprises the probability that a user in the cluster will make a purchase based on the ad.

24. The method of claim 2, wherein the method includes first initially defining the plurality of clusters.

25. The method of claim 24, wherein defining the plurality of clusters comprises utilizing user information obtained without monitoring.

26. The method of claim 24, wherein utilizing user information obtained without monitoring comprises utilizing a category tag (e.g., page group) of the page on which the item is to be displayed.

27. The method of claim 25, wherein utilizing user information obtained without monitoring comprises utilizing user information obtained from the user via a questionnaire.

28. The method of claim 24, wherein defining the plurality of clusters comprises utilizing a preexisting plurality of groups as the plurality of clusters.

29. The method of claim 24, wherein defining the plurality of clusters comprises utilizing a Bayesian network.

30. The method of claim 24, wherein defining the plurality of clusters comprises utilizing a naïve-Bayes-network clustering approach.

31. The method of claim 30, wherein utilizing a Bayesian network clustering approach comprises utilizing a bottleneck architecture.

32. The method of claim 30, wherein utilizing a Bayesian network clustering approach comprises utilizing a bottleneck architecture recursively to construct a hierarchy of clusters

33. The method of claim 30, wherein utilizing a Bayesian network clustering approach comprises training a Bayesian network using a stochastic gradient descent technique.

34. The method of claim 30, wherein utilizing a Bayesian network clustering approach comprises employing a single hidden variable having a plurality of values.

35. The method of claim 30, wherein utilizing a Bayesian network clustering approach comprises employing a plurality of hidden variables, each having two values.

36. A computer-implemented method comprising:

defining a plurality of clusters, each cluster corresponding to a group of users who are most receptive to a given type of ad; and,

allocating an ad having a particular type to at least one cluster based on the particular type of the ad and based on a predetermined criterion.

37. The method of claim 36, wherein defining the plurality of clusters comprises utilizing user information obtained without monitoring.

38. The method of claim 37, wherein utilizing user information obtained without monitoring comprises utilizing user information obtained from the user via a questionnaire.

39. The method of claim 36, wherein defining a plurality of clusters comprises defining the plurality of clusters comprises utilizing a Bayesian network.

40. The method of claim 36, wherein defining the plurality of clusters comprises utilizing a naïve-Bayes-network clustering approach.

41. The method of claim 40, wherein utilizing a Bayesian network clustering approach comprises utilizing a bottleneck architecture.

42. The method of claim 40, wherein utilizing a Bayesian network clustering approach comprises utilizing a hierarchical bottleneck architecture.

43. The method of claim 40, wherein utilizing a Bayesian network clustering approach comprises training a Bayesian network using a stochastic gradient descent technique.

44. The method of claim 40, wherein utilizing a Bayesian network clustering approach comprises employing a single hidden variable having a plurality of values.

45. The method of claim 40, wherein utilizing a Bayesian network clustering approach comprises employing a plurality of hidden variables, each having two values.

46. A computer-implemented method comprising:

determining an allocation for each of a plurality of ads to at least one of a plurality of clusters, given a constraint $\sum_j x_{ij} = q_i$, where q_i comprises a quota for ad i, and x_{ij} comprises a total number of times ad i is shown in cluster j; and,

outputting the allocation of each ad to at least one of the plurality of clusters.

47. The method of claim 46, wherein determining an allocation for each of a plurality of ads to at least one of the plurality of clusters comprises maximizing an expression $\sum_{ij} p_{ij} x_{ij}$, where p_{ij} comprises a probability that a user in cluster j will actuate ad i, given the constraint.

48. The method of claim 46, wherein determining an allocation for each of a plurality of ads to at least one of the plurality of clusters comprises determining the allocation for each of the plurality of ads to at least one of the plurality of cluster further given a constraint $\sum_i x_{ij} = c_j$, where c_j comprises a constraint for cluster j, and x_{ij} comprises a total number of times ad i is shown in cluster j.

49. The method of claim 46, further comprising:

selecting an ad for a current cluster from the allocation of each ad to the current

cluster; and,

displaying the ad.

50. A computerized system comprising:

a database storing a plurality of ads, each ad having a quota;

an allocator to allocate each of the plurality of ads to at least one of a plurality of clusters, based on a predetermined criterion accounting for at least the quota for each ad and a constraint for each cluster; and,

a communicator to select an ad for a current cluster from ads allocated to the current cluster and output the ad to a user.

51. The system of claim 50, wherein at least one of the allocator and the communicator comprises a computer program executed from a computer-readable medium by a processor.

52. The system of claim 50, wherein the database is stored as data on a computer-readable medium.

53. A machine-readable medium having instructions stored thereon for execution by a processor to perform a method comprising:

allocating each of a plurality of ads to at least one of a plurality of clusters, based on a predetermined criterion accounting for at least a quota for each ad and a constraint for each cluster;

selecting an ad for a current cluster from ads allocated to the current cluster; and, displaying the ad.

54. The medium of claim 53, wherein the predetermined criterion comprises maximizing an expression $\sum_{ij} p_{ij} x_{ij}$, where p_{ij} comprises a probability that a user in cluster j will actuate ad i.

55. The medium of claim 54, wherein the predetermined criterion further comprises maximizing the expression subject to a constraint $\sum_{j} x_{ij} = q_i$, where q_i comprises a quota for ad i, and x_{ij} comprises a total number of times ad i is shown in cluster j.

56. The medium of claim 54, wherein the predetermined criterion further comprises maximizing the expression subject to a constraint $\sum_{i} x_{ij} = c_j$, where c_j comprises a constraint for cluster j, and x_{ij} comprises a total number of times ad i is shown in cluster j.

57. The medium of claim 53, wherein allocating each of a plurality of ads to at least one of the plurality of clusters comprises determining for each ad in each cluster a probability that a user in the cluster will actuate the ad.

58. The medium of claim 53, wherein the predetermined criterion comprises maximizing an expected number of actuations of the plurality of ads, given the quota for each ad and the constraint for each cluster.

59. A machine-readable medium having instructions stored thereon for execution by a processor to perform a method comprising:

determining an allocation for each of a plurality of ads to at least one of a plurality of clusters, given a constraint $\sum_j x_{ij} = q_i$, where q_i comprises a quota for ad i, and x_{ij} comprises a total number of times ad i is shown in cluster j; and,

outputting the allocation of each ad to at least one of the plurality of clusters.

60. The medium of claim 59, wherein determining an allocation for each of a plurality of ads to at least one of the plurality of clusters comprises maximizing an expression $\sum_{ij} p_{ij} x_{ij}$, where p_{ij} comprises a probability that a user in cluster j will actuate ad i, given the constraint.

61. The medium of claim 59, wherein determining an allocation for each of a plurality of ads to at least one of the plurality of clusters comprises determining the allocation for each of the plurality of ads to at least one of the plurality of cluster further given a constraint $\sum_i x_{ij} = c_j$, where c_j comprises a constraint for cluster j, and x_{ij} comprises a total number of times ad i is shown in cluster j.

62. A computer-implemented method comprising:

applying each of at least one first item to an ordered set of rules, each rule accounting for at least a quota for each of a plurality of second items, to determine a second item for

each of the at least one first item; and,

effecting the second item for each of the at least one first item.

63. The method of claim 62, wherein each first item comprises at least information about a user, and a web page currently being browsed by the user.

64. The method of claim 62, wherein the plurality of second items comprises a plurality of ads, and effecting the second item comprises displaying the ad.

65. The method of claim 62, further initially comprising generating the ordered set of rules based on training data.

66. The method of claim 65, wherein generating the ordered set of rules comprises:

determining at least one significant correlation between a plurality of binary features of the training data and a plurality of activations of second items of the training data;

determining a second item and at least one binary feature providing a largest activation; and,

generating a rule based on the second item and the at least one binary feature providing the largest activation.

67. The method of claim 66, wherein generating the ordered set of rules further comprises:

removing records from the training data matching the rule generated; and,

repeating to generate another, lower-ordered rule while at least one significant correlation still exists.

68. The method of claim 66, wherein determining at least one significant correlation comprises utilizing one of: Chi-squared method, Fisher exact test method, and Bayesian model selection method.

69. A computer-implemented method comprising:

determining at least one significant correlation between a plurality of binary features of the training data and a plurality of activation of items from training data;

determining an ad and at least one binary feature providing a largest activation, each rule accounting for at least a quota for the item;

generating a rule based on the ad and the at least one binary feature providing the largest activation;

removing records from the training data matching the rule generated; and,

repeating to generate another, lower-ordered rule while at least one significant correlation still exists.

70. The method of claim 69, wherein each item comprises an ad.

71. A machine-readable medium having instructions stored thereon for execution by a processor to perform a method comprising:

applying each of at least one first item to an ordered set of rules, each rule accounting for at least a quota for each of a plurality of second items, to determine a second item for

each of the at least one first item; and,

effecting the second item for each of the at least one first item.

72. The medium of claim 71, the method further initially comprising generating the ordered set of rules based on training data.

73. The medium of claim 71, wherein each first item comprises at least information about a user, and a web page currently being browsed by the user, and each second item comprises an ad.

74. The medium of claim 71, wherein generating the ordered set of rules comprises:

determining at least one significant correlation between a plurality of binary features of the training data and a plurality of activations of second items of the training data;

determining a second item and at least one binary feature providing a largest activation;

generating a rule based on the second item and the at least one binary feature providing the largest activation; and,

removing records from the training data matching the rule generated; and,

repeating to generate another, lower-ordered rule while at least one significant correlation still exists.

75. A machine-readable medium having instructions stored thereon for execution by a processor to perform a method comprising:

determining at least one significant correlation between a plurality of binary features

of the training data and a plurality of activations of items from training data;

determining an ad and at least one binary feature providing a largest activation, each rule accounting for at least a quota for the item;

generating a rule based on the ad and the at least one binary feature providing the largest activation;

removing records from the training data matching the rule generated; and,

repeating to generate another, lower-ordered rule while at least one significant correlation still exists.

76. The medium of claim 75, wherein each item comprises an ad.